ÉCHANTILLONNAGE ET FLUCTUATION

1. Deux situations

Échantillonage On *connaît* la proportion p du caractère dans la population OU on *fait une hypothèse sur la valeur de p* (on est dans le cas d'une *prise de décision*) $\Rightarrow$ Utilisation d'un **intervalle de fluctuation**	**Estimation** On *ignore* la valeur de proportion p dans la population ET on ne formule pas d'hypothèse sur cette valeur. $\Rightarrow$ Utilisation d'un **intervalle de confiance**

2. Échantillonnage

2.1 Intervalle de fluctuation asymptotique

► p 442 n° 6 : fréquence
► p 442 n° 9 : fréquence - binomiale
► p 448 n° 37 : fréquence - binomiale
► p 442 n° 12 : intervalles

Démonstration

Hypothèses

- X_n est une variable aléatoire qui suit la loi binomiale $\mathscr{B}(n;p)$; et on pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$.
- X est une variable aléatoire qui suit la loi normale centrée- réduite $\mathscr{N}(0;1)$
- α un réel tel que $0 < \alpha < 1$

Démonstration

a) d'après le théorème de Moivre-Laplace, ...

...., la probabilité $P(a \leqslant Z_n \leqslant b)$ peut être approximée par la probabilité $P(a \leqslant X \leqslant b)$, c'est à dire :

....

...

b) $X \sim \mathscr{N}(0;1)$, on sait qu'il existe un unique réel u_α tel que :
$P(-u_\alpha \leqslant X \leqslant u_\alpha) = 1 - \alpha$
c'est à dire : $\displaystyle\int_{-u_\alpha}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = 1 - \alpha$

c) Pour $a = -u_\alpha$ et $b = u_\alpha$ on a donc : ....

....

d) Or $-u_\alpha \leqslant Z_n \leqslant u_\alpha$

e) En posant $\mathrm{I}_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$

$\lim\limits_{n\to+\infty} \mathrm{P}(-u_\alpha \leqslant \mathrm{Z}_n \leqslant u_\alpha) = \lim\limits_{n\to+\infty} \mathrm{P}\left(\dfrac{\mathrm{X}_n}{n} \in \mathrm{I}_n\right) = 1-\alpha$

....

....

....

....

....

....

2.2 Propriété

On sait que si X est une v.a. suivant $\mathscr{N}(0;1)$, pour tout $\alpha \in]0;1[$, il existe un unique u_α tel que $\mathrm{P}(-u_\alpha \leqslant \mathrm{X} \leqslant u_\alpha) = 1-\alpha$.

On a vu que pour $\alpha = 0{,}05$; ....

D'où : l'intervalle de fluctuation asymptotique au seuil de 95% pour une variable aléatoire X_n suivant une loi binomiale $\mathscr{B}(n;p)$ est l'intervalle :

....

....

Remarque : on ne connaît pas les valeurs p_n des probabilités de l'événement « $\dfrac{\mathrm{X}_n}{n} \in \mathrm{I}_n$ », on sait seulement que sous les conditions

....

la suite (p_n) converge vers ....

2.3 Prise de décision

- Si la fréquence observée appartient à l'intervalle de fluctuation, on accepte l'hypothèse faite sur p MAIS le risque d'erreur n'est pas quantifié !
- Si la fréquence observée n'appartient pas à l'intervalle de fluctuation, on rejette l'hypothèse faite sur p avec un risque de 5% de se tromper.

exemple : Odyssée, p 434

2.4 Applications au lycée

On réalise 4 000 lancers d'une pièce et on obtient 1 938 fois pile.
Peut-on considérer que la pièce est équilibrée ?

Ici on *fait l'hypothèse* que $p = 0{,}5$

La fréquence observée est $f =$ ....

En seconde

Pour $n \geqslant 25$ et $0{,}2 < p < 0{,}8$, l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ constitue **un** intervalle de fluctuation au seuil de 95%.

Ici $p =$ .... ,

on sait que $n =$ .... ,

donc $\mathrm{I}_1 =$ ....

On a $f \in \mathrm{I}_1$, donc ....

....

En Première

La v.a. qui compte le nombre X de pile obtenus suit la loi binomiale $\mathscr{B}(4\,000; 0{,}5)$.

« **Le** » intervalle de fluctuation au seuil de 95% associé à X est $\mathrm{I}_2 = \left[\frac{a}{n} ; \frac{b}{n}\right]$

où a est ....

....

et b est ....

....

Il n'y a pas de condition sur les valeurs de n et p.

On trouve $a =$ .... et $b =$ .... ,

donc $\mathrm{I}_2 =$ ....

On veut un risque de 5 %, on cherche les « 95 % du milieu »

k	$P(X = k)$	$P(X \leqslant k)$
1 935	0,001 526	0,020 685
1 936	0,001 627	0,022 312
1 937	0,001 734	0,024 047
1 938	0,001 846	0,025 893
1 939	0,001 963	0,027 856
...		
2 060	0,002 086	0,972 144
2 061	0,001 963	0,974 107
2 062	0,001 846	0,975 953
2 063	0,001 734	0,977 688
2 064	0,001 627	0,979 315

On a $f \in I_2$, donc ....

....

En Terminale

$n =$ .... ,

$np =$ .... ,

$n(1-p) =$ ....

Les conditions d'utilisation de l'intervalle asymptotique sont donc vérifiées. L'intervalle de fluctuation asymptotique au seuil de 0,95% est donc :

$I_3 =$ ....

....

On a .... , donc ....

....

▶ p443 n° 15 : int. fluct. / prise de décision
▶ p443 n° 16 (1 et 2) : int. fluct. / prise de décision
▶ p443 n° 21 : int. fluct. / prise de décision

Dans la vraie vie

The mystery of the missing boys
Article du 11 avril 2007 : Entre 1999 et 2003, 132 enfants sont nés dans la réserve indienne d'Aamjiwnaag, située au Canada. Sur ces 132 naissances, il y a eu 46 garçons.

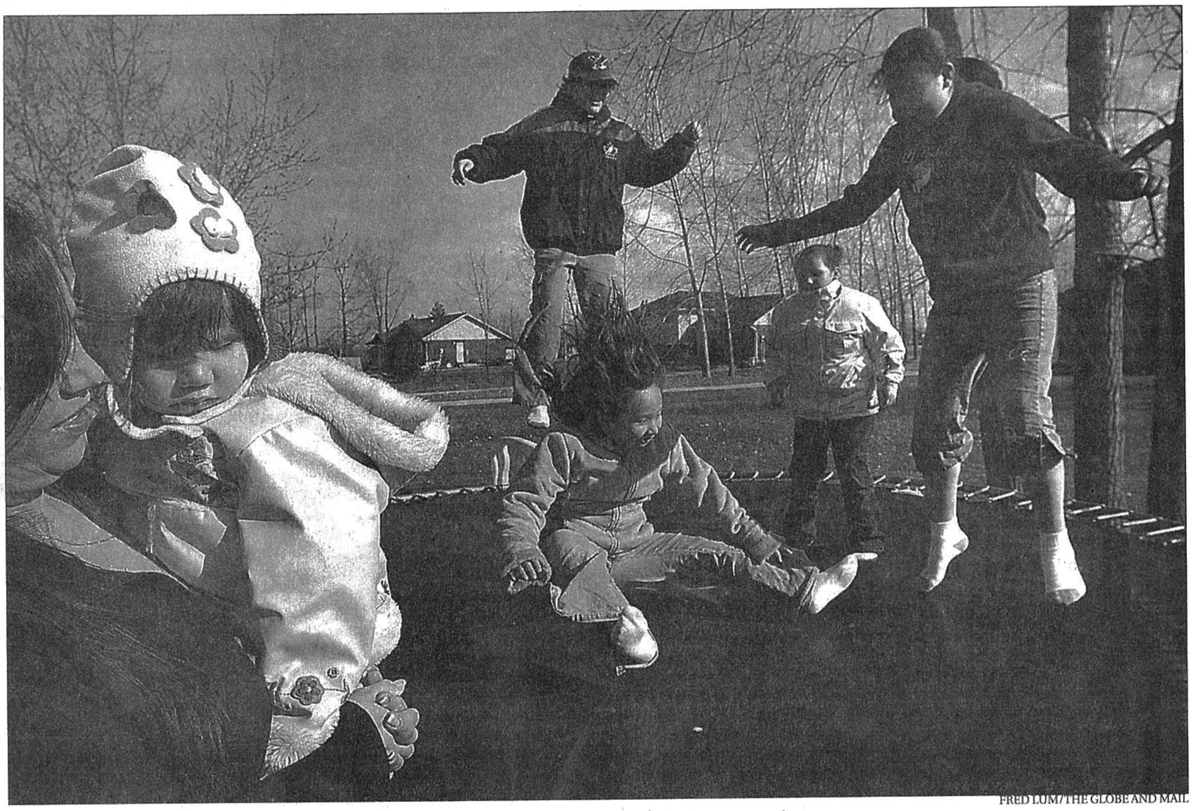
Au BAC

Amérique du Nord - mai 2013
Exercice 3 - Partie B

Les méthodes de production ont été modifiées dans le but d'obtenir 96 % de pains commercialisables.

Afin d'évaluer l'efficacité de ces modifications, on effectue un contrôle qualité sur un échantillon de 300 pains fabriqués.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la proportion de pains commercialisables dans un échantillon de taille 300.



FRED LUM/THE GLOBE AND MAIL

Lisa Joseph stands with her five children outside their home on the Aamjiwnaang First Nation Reserve near Sarnia, Ont., yesterday. Surrounded by petrochemical plants, the reserve is said to have the world's most skewed sex ratio, with nearly twice as many girls being born as boys.

ENVIRONMENTAL HEALTH

The mystery of the missing boys

Chemical pollutants flagged in new study as possible factor in skewed sex ratio

BY MARTIN MITTELSTAEDT
ENVIRONMENT REPORTER

Where are all the missing boys?

It is a question posed by a new study that has found the proportion of boys born over the past three decades has unexpectedly dropped in both the United States and Japan. In all, more than a quarter of a million boys are missing, compared to what would have been expected had the sex ratio existing in 1970 remained unchanged.

The study also says the world's most skewed sex ratio is in Canada, in a native community surrounded by petrochemical plants in Sarnia, Ont., where the number of boys born has plunged since the mid-1990s at a rate never seen.

Although the researchers do not know why boys are taking a hit, they suspect contributing causes could include widespread exposure to hormone-mimicking pollutants by women during pregnancy and by men before they help conceive children.

"We hypothesize that the decline in sex ratio in industrial countries may be due, in part, to prenatal exposure to metalloestrogens and other endocrine disrupting chemicals," said the study, issued this week in Environmental Health Perspectives, a peer reviewed journal of the U.S. National Institute of Environmental Health Sciences.

These types of chemicals include some pesticides, dioxin and methylmercury, a pollutant from coal-fired power plants and many industrial sources commonly found in seafood.

The study also flagged a host of other possible factors, including rising obesity rates, older parental age, growing stress levels, and the increasing number of children being conceived using fertility aides. Other research has shown some associations between these factors and a drop in boy births.

The study was conducted by researchers in both the U.S. and Japan, and led by Devra Lee Davis, a prominent epidemiologist and director of the Center for Environmental Oncology at the University of Pittsburgh Cancer Institute.

In an interview, Dr. Davis said that although the cause of the decline isn't known, it could be linked to the increasing number of other male reproductive problems, such as falling sperm counts and rising testicular cancer rates.

She said that males during fetal development may be more sensitive to pollutants that mimic hormones, leading to increased fetal deaths and reproductive problems later for the surviving males.

The situation in Sarnia, where nearly twice as many girls are being born than boys on the Aamjiwnaang First Nation, is internationally significant, according to the study. "To our knowledge, this is a more significantly reduced sex ratio and greater rate of change than has been reported previously anywhere," it said.

The reserve is located in the heart of Sarnia's chemical valley, and the native community, along with researchers at the University of Rochester and the Occupational Health Clinics for Ontario Workers, are trying to find the cause of the unusual sex ratio.

Fewer boys than expected are being born in the non-aboriginal community downwind of the petrochemical plants in the area, but not to the same degree as on the reserve. The work force in Sarnia has not been studied, something that would shed light on whether pollutants are the cause.

Researchers in many countries have been reporting a drop in the ratio of boys to girls being born over the past few decades.

It is considered normal in a large population for the number of baby boys to slightly outnumber girls, by a proportion of about 105 males to 100 females. It is widely thought that more boy births are a way nature compensates for higher rates of male mortality.

But the ratio has not been static in industrialized countries, and researchers suspect that increasing numbers of male fetuses are being miscarried, a kind of sex-based culling in the womb.

In Japan, the sex ratio fluctuated with no trend from 1949 to 1970, but then declined steadily to 1999, the end of the study period there.

The decline in the number of boys in Japan equals 37 out of every 10,000 births.

In the U.S., the sex ratio also declined from 1970 to 2002. The drop in the number of boys equals 17 out of every 10,000 births.

The U.S. change was concentrated among whites. There was almost no change among blacks.

le site de l'association : http://www.aamjiwnaangenvironment.ca et http://aamjiwnaangsolidarity.com
l'article : http://www.aamjiwnaangenvironment.ca/NewsArchive/MysteryoftheMissingBoys041107.pdf
une étude : http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1281269

2. Parmi les 300 pains de l'échantillon, 283 sont commercialisables.
Au regard de l'intervalle de fluctuation obtenu à la question 1, peut-on décider que l'objectif a été atteint ?

Polynésie - juin 2013
Exercice 3 - Partie 2

Thomas possède un lecteur MP3 sur lequel il a stocké plusieurs milliers de morceaux musicaux.

L'ensemble des morceaux musicaux qu'il possède se divise en trois genres distincts selon la répartition suivante :

30 % de musique classique, 45 % de variété, le reste étant du jazz.

Pendant un long trajet en train, Thomas écoute, en utilisant la fonction « lecture aléatoire » de son MP3, 60 morceaux de musique.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil 95 % de la proportion de morceaux de musique classique dans un échantillon de taille 60.
2. Thomas a comptabilisé qu'il avait écouté 12 morceaux de musique classique pendant son voyage. Peut-on penser que la fonction « lecture aléatoire » du lecteur MP3 de Thomas est défectueuse ?

3. Estimation - Intervalle de confiance

Soit X_n une v.a. suivant la loi binomiale $\mathscr{B} = (n; p)$ (p est *la proportion inconnue* du caractère).
$F_n = \dfrac{X_n}{n}$ est la fréquence associée à X_n.

exemple : Odyssée, p 434

Alors pour n suffisamment grand,

$$P\left(p \in \left[F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}}\right]\right) \geqslant 0{,}95$$

L'intervalle $\left[F_n - \dfrac{1}{\sqrt{n}}; F_n + \dfrac{1}{\sqrt{n}}\right]$ est l'intervalle de confiance de p au niveau de confiance 0,95.